Similarity matrix analysis and divergence measures for statistical detection of unknown deterministic signals hidden in additive noise

Olivier Le Bot, Jerome I. Mars, Cedric Gervaise

Similarity matrix analysis and divergence measures for statistical detection of unknown deterministic signals hidden in additive noise

O. Le Bot[1,*], J. I. Mars[*]

Univ. Grenoble Alpes, GIPSA-Lab, 11 rue des Mathématiques, Grenoble Campus BP46, F-38000 Grenoble, France

C. Gervaise[2,*]

Chaire CHORUS, Foundation of Grenoble Institute of Technology, 46 Avenue Félix Viallet, 38031 Grenoble Cedex 1, France

Abstract

This Letter proposes an algorithm to detect an unknown deterministic signal hidden in additive white Gaussian noise. The detector is based on recurrence analysis. It compares the distribution of the similarity matrix coefficients of the measured signal with an analytic expression of the distribution expected in the noise-only case. This comparison is achieved using divergence measures. Performance analysis based on the receiver operating characteristics shows that the proposed detector outperforms the energy detector, giving a probability of detection 10% to 50% higher, and has a similar performance to that of a sub-optimal filter detector.

[*]Corresponding author

Email address: lebotol@gmail.com, phone:+1 418 723 1986 -1389 (O. Le Bot)

[1]also at: Pôle STIC, ENSTA Bretagne (Université Européenne de Bretagne), 2 rue Francois Verny, 29806 Brest Cedex 9, France

[2]also at: Univ. Grenoble Alpes, GIPSA-Lab, 11 rue des Mathématiques, Grenoble Campus BP46, F-38000 Grenoble, France

Keywords: similarity matrix, divergence measures, signal detection, Receiver Operating Characteristic (ROC)
PACS: 02.50.Fz, 02.50.Ng, 05.40.-a, 05.45.Tp

1. Introduction

Deciding whether a measured data sequence is noise only or contains a short deterministic fraction within the observation time is of greatest importance in several application fields, such as radar interception, underwater acoustic signal detection, and analysis of medical signals. The general framework of a signal detector is classical, as the detector has to choose between one of the following hypotheses:

- H_0: the measured signal is noise only: x(t)=n(t)
- H_1: the measured signal has a deterministic part hidden in additive noise: x(t)=s(t)+n(t)

where n(t) is white Gaussian noise (WGN), and s(t) is the deterministic signal to be detected. To solve this signal detection, a statistical test is computed on the data that are measured, and then compared to a detection threshold [1].

The choice of the statistical test and the estimation of its probability density functions (PDFs) under hypotheses H_0 and H_1 depend on the amount of *a-priori* knowledge we have about the signal we want to detect and about the noise that it contains. When the waveform of the signal to detect is fully known, the optimum statistical test is known as a matched filter [1]. For the opposite situation, when the waveform of the deterministic signal is

not known, classical detectors are usually based on signal energy [1] or on high-order statistics [2, 3], and perform non-Gaussianity tests. Also, there are several approaches that can be used to set the detection threshold, including the Neyman-Pearson method, the Bayes' criterion, the maximum *a posteriori*, and the false discovery rate [1].

This Letter aims to present a new detection scheme using an approach that was inspired by recurrence plots [4] and is combined with divergence measures, to detect short (few tens to hundreds of samples) unknown deterministic signals in additive WGN. Recurrence plots were introduced to study the stationarity of non-linear dynamical systems [4], and have been shown to be useful for a large set of applications, like geology [5], climatology [6], music [7] and analysis of medical signals [8], to name but a few. As recurrence plots show different patterns that depend on the dynamic of the system (*i.e.*, random, periodic, chaotic), several approaches have been presented in the literature to quantify and distinguish between these three different dynamical behaviors, and particularly for deterministic signals in random process [9–15]. A common point to all of these recurrence plot studies is their use of what is known as recurrence quantification analysis (RQA) [8, 16, 17] to decide whether the measured signal is noise or not. Thus, a classical detection scheme in the recurrence plot community can be summed-up as follows:

$$x(t) \longrightarrow SM \longrightarrow RP \longrightarrow RQA \longrightarrow Detector \tag{1}$$

where SM represents the similarity matrix, and RP the recurrence plot. However, distributions of RQA metrics under hypotheses H_0 and H_1 do not generally follow existing distributions, and finding analytic expressions for these latter is not straightforward [15].

Instead of using RQA, we restrict our detector to only the use of the similarity matrix, which is sometimes called the *distance matrix* or *distance plot* in the literature [15]. The similarity matrix is the intermediate matrix that is obtained before applying the recurrence threshold that leads to the recurrence plot. Thus, we avoid the choice of this recurrence threshold and our detection scheme comes down to:

$$x(t) \longrightarrow SM \longrightarrow Detector \tag{2}$$

Our detector compares the empirical distribution of the similarity matrix coefficients of a measured signal with the distribution that is expected if the measured signal is WGN. The expression of this expected distribution can be derived analytically more easily than the RQA distribution. The comparison between the empirical and the analytic distributions is carried out with a goodness-of-fit test that is based on statistical divergences [18].

Overall, the detector presented in this Letter follows the same scheme as that proposed by Michalowicz [19]. Our algorithm differs from that of Michalowicz [19] in the use of divergence measures instead of a modified version of the χ^2 test to compare the analytic and the empirical distributions of the similarity matrix coefficients. Classical χ^2 test cannot be used because the coefficients of the similarity matrix are not fully independent of each other, as demonstrated by Michalowicz [19], which can bias the result of the test by giving much more false-positive detection than expected [19]. Finally, we do not compute the similarity matrix with a Euclidean norm only, as we propose the use of Pearson's correlation coefficient and the dot-product for this purpose [20].

After a brief recall of the recurrence plot method, we describe the different steps of our detection algorithm. Strong emphasis is put on derivation of the analytic distributions of the similarity matrix coefficients under hypothesis H_0, when the Euclidean norm, Pearson's correlation coefficient, and the dot-product are used to compute the similarity matrix. Then, we discuss the choice of an appropriate divergence function to compare the analytic and empirical distributions. The third part presents the performances of our detector through the use of receiver operating characteristic (ROC) curves. Three different deterministic signals are used in this part: a periodic signal, a chaotic Rössler system and a real acoustic signal. The influence of the degrees of freedom involved in our detection scheme are also investigated, such as the choice of the similarity function or the divergence measure. The performance of the proposed detector is compared with that of an energy detector, a sub-optimal filter detector and the optimal matched-filter detector, which are commonly used in signal processing.

2. Recurrence plots

Recurrence plots were introduced to study complex systems and are aimed at visualizing the recurrence of their phase space trajectory [4]. Transforming a data sequence to a recurrence plot representation involves three main steps.

First, the phase space trajectory of the measured signal $x(i)$ (i =1, ..., N) is reconstructed using the time delay embedding method [21, 22]. Each phase space vector is given by:

$$\overrightarrow{x_m(i)} = [x(i),\ x(i+\tau),\ \ldots\ ,\ x(i+(m-1)\tau)] \tag{3}$$

where m is the embedding dimension, and τ is the delay.

The second step consists of measuring the level of similarity between two vectors of the phase space trajectory: $\overrightarrow{x_m(i)}$ and $\overrightarrow{x_m(j)}$. Calculating the similarity between all of the possible pairs of phase space vectors leads to the similarity matrix that is defined by:

$$d(i,j) = Sim(\ \overrightarrow{x_m(i)}\ ,\ \overrightarrow{x_m(j)}\ ) \tag{4}$$

where *Sim(. , .)* is the function that is chosen to study the likeness of the phase space vectors. A lot of different mathematical functions can be used for this step. Spatial distances, and particularly the Euclidean norm, are mostly used for this purpose by the recurrence plot community [23]. In this Letter, we will introduce new functions, *i.e.*, Pearson's correlation coefficient and the dot-product, which are common similarity measures in signal processing, but not in the recurrence plot community.

Finally, as the recurrence plot is obtained through the comparison of each coefficient of the similarity matrix to a threshold, the recurrence plot is a binary matrix where the coefficient of index (i, j) is 1 if $\overrightarrow{x_m(i)}$ and $\overrightarrow{x_m(j)}$ are considered as similar, and is 0 otherwise.

3. Method

3.1. Overview of the signal detection scheme

The signal detection scheme must give an answer that allows us to decide whether a finite sequence of discrete samples contains a deterministic signal or noise only. After calculating Eq. (3) and Eq. (4), the PDF of the similarity matrix coefficients is built. This PDF is expected to fit a given theoretical PDF if the measured signal is only WGN. We use a divergence measure to

compare the theoretical expected PDF under hypothesis H_0 with the empirical PDF associated with the similarity matrix of the measured signal. We recall that in probability theory, a divergence measure is a mathematical function that quantifies the distance between two probability distributions. The result of the divergence measure is a positive number D that we compare with a detection threshold λ. If D is below this threshold, this means that the distributions look alike, and consequently that the measured signal is WGN. For the opposite, *i.e.*, if D is greater than the threshold, this means that the empirical PDF differs from the theoretical noise PDF, and thus that a deterministic signal is present. The threshold λ is chosen according to the Neyman-Pearson criterion. We recall that when performing a hypothesis test between two hypothesis H_0 versus H_1, Neyman-Pearson criterion is the one that maximizes the probability of detection while guaranteeing a given probability of false alarm (Pfa). With other words, a threshold fixed by the Neyman-Pearson criterion maximizes the probability (Pd) of choosing hypothesis H_1 when H_1 is effectively true and rejects hypothesis H_0 with a probability Pfa when H_0 is effectively true. To apply this criterion, we use Monte-Carlo simulations to built the distribution of the divergence measures D between the analytic PDF expected under hypothesis H_0 and the empirical PDF of the similarity matrix coefficients of finite length WGN. All of the steps of this detection scheme are summarized in Figure 1.

3.2. Analytical distribution of the similarity matrix coefficients in the 'noise only' case

3.2.1. Hypothesis

Under hypothesis H_0, we assume that the measured samples $x(1)$, $x(2)$, ..., $x(n)$ from a given sequence are independent Gaussian random variables with zero mean and variance σ^2.

To obtain the similarity matrix, we look at the similarity between the vectors $\overrightarrow{x_m(i)} = [x(i),\ x(i+\tau),\ \ldots,\ x(i+(m-1)\tau)]$ and $\overrightarrow{x_m(j)} = [x(j),\ x(j+\tau),\ \ldots,\ x(j+(m-1)\tau)]$ (with $i \neq j$), the components of which come from the measured signal. Therefore, under hypothesis H_0, the components of both of these vectors are also independent Gaussian random variables with zero mean and variance σ^2, and the vectors are independent of each other.

Based on these assumptions, we analytically model the PDF of the similarity matrix coefficients under hypothesis H_0, when the Euclidean norm, Pearson's correlation coefficient, and the dot-product (each of which is described below) are used to compare the state space vectors.

3.2.2. Euclidean norm

The Euclidean norm between the two state space vectors is given by:

$$d_{i,j} = \sqrt{\sum_{k=1}^{m} \left(x_k(i) - x_k(j)\right)^2} \tag{5}$$

where $x_k(i)$ and $x_k(j)$ are the k^{th} components of vectors $\overrightarrow{x_m(i)}$ and $\overrightarrow{x_m(j)}$, respectively. According to the assumptions made above, $x_k(i)$ and $x_k(j)$ are independent Gaussian random variables with zero mean and variance σ^2. Therefore, $y_k = x_k(i) - x_k(j)$ is also a Gaussian random variable with zero

mean and variance $2\sigma^2$, and every y_k is independent of every other y_k, for all k.

By definition, if we take m independent Gaussian random variables W_k with zero mean and variance σ_W^2, then the random variable Z given by

$$Z = \sum_{k=1}^{m} \left(\frac{W_k}{\sigma_W}\right)^2 \tag{6}$$

has a χ^2 distribution with m degrees of freedom. By analogy, we show that the random variable

$$\sum_{k=1}^{m} \frac{y_k^2}{2\sigma^2} = \frac{\sum_{k=1}^{m} \left(x_k(i) - x_k(j)\right)^2}{2\sigma^2} = \frac{d_{i,j}^2}{2\sigma^2} \tag{7}$$

has a χ^2 distribution with m degrees of freedom.

So, if the Euclidean norm is used, normalizing the coefficients of the similarity matrix as in Eq. (7) will give a new similarity matrix, the coefficients of which will have a χ^2 distribution with m degrees of freedom.

3.2.3. Pearson's correlation coefficient

Pearson's correlation coefficient between two state space vectors is given by

$$d_{i,j} = \frac{\sum_{k=1}^{m} \left(x_k(i) - \overline{x_k(i)}\right)\left(x_k(j) - \overline{x_k(j)}\right)}{\sqrt{\sum_{k=1}^{m} \left(x_k(i) - \overline{x_k(i)}\right)^2} \sqrt{\sum_{k=1}^{m} \left(x_k(j) - \overline{x_k(j)}\right)^2}} \tag{8}$$

where $x_k(i)$ and $x_k(j)$ are the k^{th} components of vectors $\overrightarrow{x_m(i)}$ and $\overrightarrow{x_m(j)}$, respectively, and $\overline{x_k(i)}$ and $\overline{x_k(j)}$ are the empirical means of $\overrightarrow{x_m(i)}$ and $\overrightarrow{x_m(j)}$, respectively. According to the assumptions made above, as $x_k(i)$ and $x_k(j)$ are independent variables for all k, then their co-variance is zero and the joint PDF of pair ($x_k(i)$, $x_k(j)$) is the product of their respective PDFs.

As $x_k(i)$ and $x_k(j)$ have Gaussian distributions with zero mean and variance σ^2, their joint PDF is given by:

$$f(x_k(i), x_k(j)) = \frac{1}{2\pi\sigma^2} exp\Big(- \frac{x_k(i)^2}{2\sigma^2} \Big) exp\Big(- \frac{x_k(j)^2}{2\sigma^2} \Big) \tag{9}$$

which is exactly the same as that of a bi-variate normal distribution with independent random variables and zero mean. Fisher [24–26] demonstrated that for pairs of independent random variables with bi-variate Gaussian distributions, the distribution of the Pearson's correlation coefficient r can be expressed as:

$$f(r) = \frac{1}{\beta\Big(\frac{m-1}{2}, \frac{1}{2}\Big)} (1 - r^2)^{\frac{m-4}{2}} \tag{10}$$

where $\beta(.\ ,\ .)$ is the Beta function, and m is the embedding dimension.

3.2.4. Dot-product

The dot-product between two state space vectors is given by:

$$d_{i,j} = \sum_{k=1}^{m} x_k(i) \times x_k(j) \tag{11}$$

where $x_k(i)$ and $x_k(j)$ are the k^{th} components of vectors $\overrightarrow{x_m(i)}$ and $\overrightarrow{x_m(j)}$, respectively. Eq. (11) can be rewritten as follows:

$$d_{i,j} = \frac{1}{4}\Bigg(\sum_{k=1}^{m} \Big(x_k(i) + x_k(j) \Big)^2 - \sum_{k=1}^{m} \Big(x_k(i) - x_k(j) \Big)^2 \Bigg) \tag{12}$$

such that we rely on the PDF of $\sum_{k=1}^{m} \left(x_k(i) + x_k(j) \right)^2$ and $\sum_{k=1}^{m} \left(x_k(i) - x_k(j) \right)^2$, which are easier to use, to derive the PDF associated with Eq. (11).

In the case where $\sum_{k=1}^{m} \left(x_k(i) + x_k(j) \right)^2$, as $y_k = x_k(i) + x_k(j)$ is the sum of two independent Gaussian random variables with zero mean and variance

σ^2, then y_k also has a Gaussian distribution with zero mean and variance $2\sigma^2$. As in section 3.2.2, we show that the random variable $u_{i,j}$ given by:

$$u_{i,j} = \sum_{k=1}^{m} \frac{y_k^2}{2\sigma^2} = \sum_{k=1}^{m} \frac{\left(x_k(i) + x_k(j)\right)^2}{2\sigma^2} \tag{13}$$

has a χ^2 distribution with m degrees of freedom. The same demonstration holds for $\sum_{k=1}^{m} \left(x_k(i) - x_k(j)\right)^2$, and as in section 3.2.2, the random variable $v_{i,j}$ that is given by:

$$v_{i,j} = \sum_{k=1}^{m} \frac{\left(x_k(i) - x_k(j)\right)^2}{2\sigma^2} \tag{14}$$

follows a χ^2 distribution, with m degrees of freedom.

Combining Eq. (12), (13) and (14), this leads to:

$$\frac{2d_{i,j}}{\sigma^2} = u_{i,j} - v_{i,j} \tag{15}$$

Therefore, the distribution of the dot-product of two state space vectors that satisfy our assumptions is equivalent, to a scaling factor, to the difference of two independent random variables with χ^2 distributions. The analytic expression of the distribution associated with this difference can be derived using the moment-generating function of the χ^2 distribution.

If X_1, X_2, ..., X_n are n independent random variables (which are not necessarily identically distributed), and S_n is a random variable defined by:

$$S_n = \sum_{i=1}^{n} a_i X_i \tag{16}$$

where $a_i \in \mathbb{R}$ is a constant, then the moment-generating function of S_n is given by:

$$M_{S_n}(y) = M_{X_1}(a_1 y) \times M_{X_2}(a_2 y) \times ... \times M_{X_n}(a_n y) \tag{17}$$

where M_{X_i} is the moment-generating function of X_i.

In our case, S_n is the sum of two independent random variables ($u_{i,j}$, $v_{i,j}$), both of which follow a χ^2 distribution with m degrees of freedom. The moment-generating function of a χ^2 distribution is:

$$M_X(y) = (1-2y)^{-\frac{m}{2}} \tag{18}$$

According to the properties given above, the moment-generating function of $u_{i,j} - v_{i,j}$ is therefore given by:

$$M_{u_{i,j}-v_{i,j}}(y) = (1-4y^2)^{-\frac{m}{2}} \tag{19}$$

The moment-generating function obtained in Eq. (19) is the same as that of a variance-Gamma distribution, the general expression for which is given by:

$$M_{V.G.}(\lambda,\alpha,\beta,\mu,y) = e^{\mu y}\left[\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+y)^2}\right]^{\lambda} \tag{20}$$

By identification, we find that the parameters of Eq. (20) leading to Eq. (19) are: $\mu = 0, \quad \alpha = 1/2, \quad \beta = 0, \quad \lambda = m/2$

The PDF of a variance-Gamma distribution is defined as:

$$f_Z(z) = \frac{(\alpha^2-\beta^2)^{\lambda}|z-\mu|^{\lambda-\frac{1}{2}}}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-\frac{1}{2}}} K_{\lambda-\frac{1}{2}}(\alpha|z-\mu|)e^{\beta(x-\mu)} \tag{21}$$

where Γ is the gamma function, and $K_\nu(x)$ is the modified Bessel function of second kind. The PDF of $(u_{i,j} - v_{i,j})$ is finally obtained by replacing μ, α, β and λ by the values defined above, which gives:

$$f_{(u_{i,j}-v_{i,j})}(z) = \frac{|z|^{\frac{m-1}{2}}}{2^m\sqrt{\pi}\Gamma\left(\frac{m}{2}\right)} K_{\frac{m-1}{2}}\left(\frac{|z|}{2}\right) \tag{22}$$

This PDF is continuous when $z = 0$ and $m > 1$, and is given by

$$\lim_{z \to 0} f_{(u_{i,j} - v_{i,j})}(z) = \frac{1}{4\sqrt{\pi}} \frac{\Gamma\left(\frac{m-1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} \tag{23}$$

So, if the dot-product is used, the multiplication of the similarity matrix by a factor $2/\sigma^2$ will give a new similarity matrix, the coefficients of which will have a distribution that is defined by Eq. (22) and (23).

3.3. The divergence measure between the analytic distribution for the 'noise only' case, and the empirical distribution of an unknown signal

The next step in our detection scheme (fourth block in Fig. 1) is the comparison between the analytic distributions defined in the previous section and the empirical distributions of the similarity matrix coefficients of an unknown signal, to decide whether this latter fits the expected distribution under hypothesis H_0. Such a comparison is called a goodness-of-fit test in statistics. A popular goodness-of-fit method is Pearson's χ^2 test [27]. As stated before, in our case, the χ^2 test gives more false positives than expected, which means that it is not usable.

To avoid this problem, we propose to use a goodness-of-fit test that is based on an *information theory* approach, which consists of computation of the difference of entropy between the two PDFs we want to compare. This approach is called the divergence measure, and this was first proposed by Shannon [28] and Kullback [29].

3.3.1. Divergence measures

Mathematically speaking, a function $Div(.\ ,\ .) : X \times X \mapsto \mathbb{R}$ (where X is a set) is a divergence function [30] if, for all $x,\ y \in X$, it has the following

properties:

- $Div(x,\ y) \geq 0$ (non-negativity);
- $Div(x,\ y) = 0 \iff$ x=y (identity of indiscernible).

In general, divergence measures do not satisfy the triangular inequality; some of them are symmetric. Divergence measures can be split into several classes; *i.e.*, f-divergences, Bregman divergences, α-divergences, β-divergences, and γ-divergences. Thorough state-of-the-art reviews of divergence classes and their respective properties can be found in [18, 30, 31].

It is essential to decide which divergence measure to use, to have the best processing gain for our detector, although to the best of our knowledge, there are no strict rules for this in the literature. Therefore, we decided to study the performances of our detector for three of the most-cited divergence measures in the literature: the Kullback-Leibler divergence, the Hellinger divergence, and the Jensen-Shannon divergence.

The Kullback-Leibler divergence is probably one of the most used divergences in the literature, particularly for goodnes-of-fit tests and parametric estimations [32, 33]. The Kullback-Leibler divergence is defined as follows:

$$D_{KL}(p \mid\mid q) = \int p(x) ln\left(\frac{p(x)}{q(x)}\right) dx \tag{24}$$

where *p(x)* and *q(x)* are the PDFs to be compared.

The Hellinger divergence is expressed as:

$$D_H(p \mid\mid q) = \sqrt{\frac{1}{2}\int \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx} \tag{25}$$

The Kullback-Leibler divergence and Hellinger divergence belong to the f-divergence class.

The last divergence that we investigate here is the Jensen-Shannon divergence, which can be seen as a symmetric and smoothed version of the Kullback-Leibler divergence, and is expressed as:

$$D_{JS}(p||q) = \eta D_{KL}\Big(p \;||\; (\eta p + (1-\eta)q)\Big) + \eta D_{KL}\Big(q \;||\; (\eta p + (1-\eta)q)\Big) \tag{26}$$

where $D_{KL}(.||.)$ is the Kullback-Leibler divergence given by Eq. (24), and $\eta \in [0,\ 1]$. In this Letter, the simulations are carried out with $\eta = 0.5$ (as an arbitrary choice).

3.3.2. Distributions of the divergences in the 'noise only' case

To decide whether a measured signal is noise only or is a deterministic signal, the result of each of these divergence measures is compared to a threshold λ, the value of which is chosen to guarantee a given Pfa. To achieve this, it is necessary to know the PDF of the divergence measures under hypothesis H_0. We obtain this latter with Monte-Carlo simulations (with 50,000 repetitions), by generating WGN (zero mean, $\sigma^2 = 1$), and computing the first four steps of our detection scheme with an embedding $m = 16$ and a delay $\tau = 1$ (Fig. 1). We repeat this simulation to obtain 50,000 values of the divergence measures, and build their PDF. This PDF is estimated by a classical histogram method. These simulations are carried out for each similarity function (*i.e.*, Euclidean norm, Pearson's correlation coefficient, dot-product), followed by the three different divergence measures presented above, which leads to nine different detectors.

4. Results

The performances of these nine detectors are studied through the ROC curves, which display the probability of detection *versus* the Pfa associated with the detector, as a function of the detection threshold. Within this performance analysis section, we show that the Kullback-Leibler divergence always gives the best processing gain whatever the similarity function (*i.e.*, Euclidean norm, Pearson's correlation coefficient, or dot-product) used to built the similarity matrix. Then, we establish which similarity function (*i.e.*, Euclidean norm, Pearson's correlation coefficient, or dot-product) should be associated with the Kullback-Leibler divergence to give the best overall performances. Finally, we compare the performances of our detector with the energy detector, a sub-optimal filter detector and the optimal matched-filter detector.

4.1. Performance analysis methodology

To build the ROC curves, it is necessary to know the distribution of the divergence measures under hypothesis H_1. This distribution is obtained with Monte Carlo simulations of 50,000 experiments. The performances are studied through two simulated deterministic signals, namly, a periodic signal given by a cosine function (section 4.3) and the first component of a Rössler system in chaotic regime (section 4.4), as well as with a real acoustic signal (section 4.5). In a passive context, the duration of the signal to be detected is generally not known. Therefore, we study the performances where the length of the deterministic signal to detect is shorter than the observation time. The PDFs of the divergence measures under hypothesis H_1 are constructed

when the deterministic signal occupies $T\%$ of the observation time, with $T \in \{10,\ ...\ ,\ 100\}$ (Fig 2).

Several signal-to-noise ratios (SNRs) between - $2dB$ and + $4dB$ are also studied (by step of 0.5 dB). Only the most significant results are shown in this Letter. We recall that the SNR expressed in decibel is defined as:

$$SNR_{dB} = 10log_{10}\frac{\frac{1}{L_s}\sum_{i=1}^{L_s} s(i)^2}{\frac{1}{L_b}\sum_{j=1}^{L_b} b(j)^2} \tag{27}$$

where $s(i)$ is the deterministic signal and L_s is its length, $b(j)$ is the WGN and L_b is its length. Therefore, the targeted SNR is obtained by adjusting the variance of the WGN with respect to energy of the deterministic signal as follow:

$$\sigma_b^2 = \left(\frac{1}{L_s}\sum_{i=1}^{L_s} s(i)^2\right).10^{-SNR_{dB}/10} \tag{28}$$

For each simulated signal under hypothesis H_1, we compute the first four steps of our detection scheme to obtain the divergence measures (Fig. 1). The PDFs of the divergence measures under hypotheses H_0 and H_1 lead to the construction of the ROC curves, which depend on the values chosen for the pair ($SNR,\ T$).

To see where our detector is positioned relative to classical detectors from the literature, we compare its ROC curves with those of the energy detector, a sub-optimal filter detector and the optimal matched-filter detector. The latter two detectors will only be used when the deterministic signal is the periodic signal, as they can hardly be used with a chaotic signal and a real

acoustic signal, which have, *a priori*, an unknown waveform. In the next subsection, we recall the underlying PDFs of these three detectors under hypotheses H_0 and H_1.

4.2. Detectors of reference

4.2.1. Energy detector

The energy detector is commonly used in signal processing when nothing is known about the signal to be detected. This detector is based on the random variable g, which is defined as follows:

$$g = \sum_{i=1}^{L} x(t_i)^2 \tag{29}$$

where *x(t)* is the measured signal. Under hypothesis H_0, the measured signal is WGN with zero mean and variance σ^2. Therefore, the random variable g/σ^2 has a χ^2 distribution with L degrees of freedom.

For hypothesis H_1, when the deterministic signal is a cosine with length L_s, g is given by:

$$g = \sum_{i=1}^{L} \Big(b(t_i) + Acos(2\pi f_0 t_i) rect_{L_s}(t_i) \Big)^2 \tag{30}$$

with A the amplitude of the cosine, f_0 its frequency and $rect_{L_s}(t_i)$ a rectangular window of length L_s. Then, the random variable g/σ^2 has a noncentral χ^2 distribution with L degrees of freedom and a noncentrality parameter $\zeta = L_s A^2/2\sigma^2$ [34]. When the deterministic signal is the chaotic Rössler system or the real acoustic signal, the distribution of g under hypothesis H_1 is obtained empirically with Monte-Carlo simulations.

4.2.2. Sub-optimal filter detector

As stated in the Introduction, when the waveform of the signal to be detected is perfectly know, the optimum detector is called a matched filter. Here, we consider the sub-optimal case where the detector includes all of the characteristics of the cosine signal it has to detect, excepted its duration. Thus the detector expects the cosine to be present 100% of the time, whereas it will effectively be present only T% of the time. This detector is based on the random variable g, which is given by:

$$g = \sum_{i=1}^{L} x(t_i) \times Acos(2\pi f_0 t_i) \tag{31}$$

where L is the length of the measured signal $x(t)$. Under hypothesis H_0, the measured signal is WGN with zero mean and variance σ^2. We can demonstrated that the random variable g has a normal distribution with zero mean and variance $L\sigma^2A^2/2$ [1].

For hypothesis H_1, when the deterministic signal is a cosine with length L_s, g is given by:

$$g = \sum_{i=1}^{L} \Big(b(t_i) + Acos(2\pi f_0 t_i).rect_{L_s}(t_i) \Big) \times Acos(2\pi f_0 t_i) \tag{32}$$

and has a normal distribution with mean $L_sA^2/2$, and variance $L\sigma^2A^2/2$ [1].

4.2.3. Optimal Matched-filter detector

In this section we consider the optimal matched-filter detector, i.e. the detector knows all of the characteristics of the cosine signal it has to detect.

This detector is based on the random variable g, which is given by:

$$g = \sum_{i=1}^{L} x(t_i).Acos(2\pi f_0 t_i).rect_{L_s}(t_i) \tag{33}$$

where L is the length of the measured signal $x(t)$. Under hypothesis H_0, the measured signal is WGN with zero mean and variance σ^2. We can demonstrated that the random variable g has a normal distribution with zero mean and variance $L_s\sigma^2A^2/2$ [1]. For hypothesis H_1, when the deterministic signal is a cosine, g is given by:

$$g = \sum_{i=1}^{L} \Big(b(t_i) + Acos(2\pi f_0 t_i).rect_{L_s}(t_i)\Big).Acos(2\pi f_0 t_i).rect_{L_s}(t_i) \tag{34}$$

and has a normal distribution with mean $L_sA^2/2$, and variance $L_s\sigma^2A^2/2$ [1].

4.3. Performances with a periodic signal

In this section, the deterministic signal to detect is a cosine function, the frequency of which, f_0, is randomly chosen for each experiment (uniform distribution), so that $f_0/f_e \in [0.05\ \ 0.45]$, where f_e is the sampling frequency. The cosine function is added to WGN for 100 samples. Thus, the distributions of the divergence measures obtained under hypothesis H_0 in section 3.3.2 are calculated with a WGN for 100 samples.

We recall that for a cosine with amplitude A added to WGN with zero mean and variance σ^2, the SNR is $A^2/2\sigma^2$, or in dB, $10log_{10}(A^2/2\sigma^2)$. To sum-up, in this section, the measured signal under hypothesis H_1 is given by $x(t) = b(t) + Acos(2\pi f_0 t)rect_{L_s}(t)$, where *b(t)* is a WGN sequence, $rect_{L_s}(t)$ is a rectangular window of length L_s, A is the amplitude of the cosine, and f_0 is the frequency of the cosine.

All of the results in this section are given for an embedding dimension $m = 16$ and $\tau = 1$. However, all of the conclusions remain the same for other embeddings within the range [8, 20]. All of the ROC curves are identical for $m \in [12\ , 18]$.

4.3.1. Performances as a function of the divergence measure

First, we look at the influence of the divergence measure on the performances of the detector, for each similarity function taken separately.

As ROC curves aim at showing the probability of detection (Pd) as a function of the probability of false alarm (pfa), the performances of a given detector are considered as good when its Pd is close or equal to 1 whatever the value of Pfa. At the contrary, performances are considered as bad when $Pd = Pfa$. Also, a detector is considered better than another detector, if the COR curve of the first detector is above the ROC curve of the second detector.

For example, on Fig. 3, we see that for each subplot associated with a couple (T, SNR), the ROC curve of the detector using the Kullback-Leibler divergence (plain line) is above the ROC curve of the Hellinger and Jensen-Shannon divergences. Therefore, we can say that the detector using the Euclidean norm with the Kullback-Leibler divergence outperforms the detectors using the Euclidean norm with the Hellinger divergence and the Jensen-Shannon divergence (Figs. 3). The same observations and conclusions hold when the dot-product is used to compute the similarity matrix (Figs. 5). Whatever the couple (T, RSB) used for the simulations, the ROC curves of the detector using the dot-product with the Kullback-Leibler divergence are above the ROC curve with the Hellinger divergence and the Jensen-Shannon

divergence (Figs. 5). With these two similarity functions (i.e. Euclidean norm and dot-product), when the SNR is positive (three last rows of the panel) and Pfa $\leq 10^{-3}$, the detection probability with the Kullback-Leibler divergence is 10% to 50% greater than those obtained with the Hellinger divergence and Jensen-Shannon divergence.

When Pearson's correlation coefficient is used, the Kullback-Leibler divergence is slightly better than the Hellinger divergence and Jensen-Shannon divergence (Fig. 4).

According to these results, for the remainder of this section 4.3, we have chosen the Kullback-Leibler divergence to compare the analytic and empirical PDFs of the similarity matrix coefficients, whichever similarity function is used to compute the similarity matrix.

4.3.2. Performances as a function of the similarity function

We now look at the similarity function that gives the best results for the detector, when the Kullback-Leibler divergence is used to compare the analytic and empirical distributions. We compare ROC curves of the detectors having the following configurations: {Euclidean Norm, Kullback-Leibler}, {Pearson's correlation coefficient, Kullback-Leibler}, {dot-product, Kullback-Leibler} (Fig. 6).

For all of the combinations of SNR and T, the detector using the dot-product performs the best, followed by the detector with the Euclidean norm, and then last, the detector using Pearson's correlation coefficient. When the Pfa is around 10^{-4}, the detection probability of the detector using the dot-product is 10% to 25% higher than the detector with the Euclidean norm, and 10% to 80% higher than that with Pearson's correlation coefficient, which

depends on the values given to the pair (RSB, T). For a given SNR, we find for all similarity functions that the data change quickly when T increases. The same observation is made when T is constant and the SNR increases by a few decibels.

4.3.3. Comparison with the detectors of reference

The proposed detector with the dot-product and Kullback-Leibler divergence is compared with the energy detector and the matched-filter detector, in terms of their ROC curves (Fig. 7). For all of the SNR values > 0, the performances of the proposed detector are higher than those of the energy detector, whatever the length T of the cosine. For a Pfa around 10^{-4} the difference in terms of the detection probability between both of the detectors is between 0.05 and 0.45, depending on the SNR and T.

For most of the (SNR, T) combinations, the proposed detector has a similar performance to the sub-optimal filter detector. Our detector is significantly better than the sub-optimal filter detector only when the cosine is very short ($T \leq 30$ %) and has a SNR > 2 dB. However, the performances of the proposed detector are far behind those of the optimal matched-filter detector.

These performances for the proposed detector can be explained as follows. As the similarity matrix is computed by splitting the signal into several state space vectors, when hypothesis H_1 is true, some of these vectors correspond to the signal we want to detect. Therefore, our detector is self-fed by vectors associated with the useful signal and is locally equivalent to a matched-filter detector. If a state space vector $\overrightarrow{s_m(t_i)}$ that contains samples from the deterministic signal is compared to a vector $\overrightarrow{s_m(t_j)}$ that has only noise samples,

then the coefficient (i, j) of the similarity matrix belongs to the PDF associated with the noise-only case. For the opposite, if vector $\overrightarrow{s_m(t_j)}$ also contains samples from the deterministic signal, we are back under hypothesis H_1 of a classical matched filter. In the end, some of the coefficients of the similarity matrix correspond to hypothesis H_0 and follow the analytic distribution derived above in the noise-only case, while the remaining coefficients correspond to hypothesis H_1 and do not follow this analytic distribution. The empirical PDF of the coefficients of the similarity matrix differs significantly from the one that would be expected in the noise-only case, and the deterministic signal is detected, even when this latter is short and has a poor SNR.

4.4. Performances with a Rössler system in chaotic regime

In this section, the deterministic signal to detect is the first component (or x-component) of a Rössler system. This system is defined by:

$$\dot{x} = -y - z \tag{35}$$

$$\dot{y} = x + ay \tag{36}$$

$$\dot{z} = b + z(x - c) \tag{37}$$

We take $a = 0.15$, $b = 0.2$ and $c = 10$, so that it has a chaotic behavior. The sampling time Δt is equal to 0.4 s. The component x(t) is added to WGN for 200 samples, so that the Rössler system has enough time to oscillate during a few periods, even when it occupies a small percentage T of the observation time. Thus, the distributions of the divergence measures obtained under hypothesis H_0 in section 3.3.2 are now calculated with a WGN for 200 samples. The SNR is adjusted according to Eq. (28). For each realization

of the Monte-Carlo simulations, the initial conditions $[x(0),\ y(0),\ z(0)]$ are randomly chosen with uniform distribution within the range [-5 , 5] in order to get various waveform for $x(t)$.

We approach this part through two points of view. In section 4.4.1, we do not take into account that $x(t)$ is part of a 3-components system and study the performances of the detector as we did with the cosine, i.e. with $\tau = 1$ and $m \in [8, 20]$. Then, in section 4.4.2, we take into account that x(t) comes from a 3-components system and so that its phase space trajectory can be reconstructed with $m = 3$. Thus, in 4.4.2 we study the performances for $m = 3$ and $\tau \in [2, 9]$.

4.4.1. Case 1: $\tau = 1$*,* $m \in [8, 20]$

We do not show all the ROC curves as we did in the previous section, but only give a summary of the main results. All of the results are given for an embedding dimension $m = 16$. However, all of the conclusions remain the same for other embeddings within the range [8, 20].

As for the periodic signal in section 4.3, we first looked at the divergence measure giving the best detection performances, for each similarity function taken separately. Results and conclusions remain the same as for the periodic signal, namely that whatever the similarity function $Sim(.\ .)$ used to calculate the similarity matrix, the Kullback-Leibler divergence always gives the best detection capabilities. The Hellinger divergence and Jensen-Shannon divergence have much lower performances than Kullback-Leibler divergence.

Secondly, when the Kullback-Leibler divergence is used as a divergence measure, then the best overall detection performances are again obtained

with the dot-product as $Sim(.\ .)$ function to compute the distance matrix.

Finally, on Fig. 8, we compare the COR curves of the energy detector with those of the proposed method with the dot-product and Kullback-Leibler divergence. We see on this figure that whatever the couple (T, RSB) chosen, the proposed detector always outperforms the energy detector. For $Pfa < 10^{-3}$, the probability of detection of proposed detector is 20 % to 50 % higher than the one of the energy detector.

4.4.2. Case 2: $\tau \in [2, 9]$, $m = 3$

All of the results in this section are given for $m = 3$ and $\tau = 3$. However, all of the conclusions remain the same for other τ within the range [2, 9].

Like in previous sections, we found that Kullback-Leibler divergence is the divergence measure that gives the best detection performances. Then, we noticed that by associating the Kullback-Leibler divergence with the Euclidean norm, we get slightly better performances than by combining Kullback-Leibler divergence with the dot-product (see Fig. 9). The probability of detection increases only by a few percent between the dot-product and the Euclidean norm. At last, when comparing the proposed detector with the energy detector, we see that the energy detector and the proposed detector with {Euclidean norm, Kullback-Leibler divergence} give very similar results and that the energy detector is slightly better than the proposed detector with {dot-product, Kullback-Leibler divergence} (Fig. 9).

4.5. Performances with a real acoustic signal

In this section, we test the performances of the proposed detector with a

real underwater acoustic signal that was recorded in the Mediterranean Sea in August 2014. This sound, whose waveform is given on Fig. 2g, was produced by a fish. As the background noise mixed with the fish sound is not a true WGN, the necessary assumptions given in section 3.2.1 are not met and so we could not retrieve the theoretical distributions we found previously under H_0 hypothesis. Therefore, we have extracted the fish sound and add it with a simulated WGN. Like for previous simulations we change the duration of the noise so that the fish sound occupies T % of the observation time. Various SNR are also tested, according to Eq. (28).

Fig. 10 shows the COR curves of the proposed detector with $\tau = 1$, $m = 16$, the dot-product as similarity function and the Kullback-Leibler divergence as divergence measure. Like for the periodic signal and for the Rössler system, the proposed detector outperforms the energy detector for all couples (T, RSB).

5. Conclusion

This Letter has presented a scheme that is based on statistical analysis of the similarity matrix coefficients and on divergence measures to detect an unknown deterministic signal in WGN. Under hypothesis H_0, the distribution of the similarity matrix coefficients was derived analytically for three similarity functions: the Euclidean norm, Pearson's correlation coefficients, and the dot-product. Then, divergence measures were used to compare this analytic distribution with the empirical distribution of a measured signal for which we wanted to apply the detection test. Three divergence measures were tested in this study: Kullback-Leibler divergence, Hellinger divergence, and Jensen-Shannon divergence. The performance of the detector was stud-

ied through ROC curves. The influences of similarity functions, divergence measures, length of the deterministic signal, and the SNR were discussed. We found that the Kullback-Leibler divergence was always the divergence measure giving, in the end, the best results. The best overall performances are obtained when this divergence is used with the dot-product as similarity function. The proposed detector provided with the dot-product and the Kullback-Leibler divergence was compared with the energy detector, a sub-optimal filter detector and the optimal matched-filter detector. The results with two simulated deterministic signal, namely a periodic signal and a chaotic Rössler system, as well as with a real underwater acoustic signal show that the proposed detector has a much better detection probability than the energy detector and similar performance to the sub-optimal filter detector. At last, results shown in this Letter and our own experience on other real signals indicate that by performing a statistical analysis of similarity matrix coefficients we get enhanced detection performances relative to the classical energy detector, independently of the kind of the deterministic signal to be detected.

References

[1] S. M. Kay, Fundamentals of statistical signal processing, volume 2: Detection theory, Prentice Hall, 1993.

[2] A. Swami, G. Giannakis, G. Zhou, Signal Proc. 60 (1997) 65.

[3] C. Gervaise, A. Barazzutti, S. Busson, Y. Simard, N. Roy, Appl. Acoust. 71 (2010) 1144–1163.

[4] J. P. Eckmann, S. Kamphorst, D. Ruelle, Europhys. Lett. 4 (1987) 973.

[5] N. Marwan, M. Thiel, N. K. Nowaczyk, Nonlinear Processes in Geophysics 1 (2002) 325.

[6] S. Frey, F. Sadlo, T. Ertl, IEEE Trans. on Visualization and Computer Graphics 18 (2012) 2023.

[7] J. Serrà, X. Serra, R. G. Andrzejak, New J. of Phys. 11 (2009) 093017.

[8] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Phys. Rev. E 66 (2002) 026702.

[9] J. P. Zbilut, A. Giuliani, C. L. Webber, Phys. Lett. A 237 (1998) 131.

[10] J. P. Zbilut, A. Giuliani, C. L. Webber, Phys. Lett. A 246 (1998) 122.

[11] J. P. Zbilut, A. Giuliani, C. L. Webber, Phys. Lett. A 267 (2000) 174.

[12] B. Dissinger, G. Rohde, R. Rhodes Jr., F. Bucholtz, J. Nichols, Intensity analysis of recurrence plots for the detection of deterministic signals in noise, Tech. rep., NRL (2006).

[13] T. Aparicio, E. F. Pozo, D. Saura, J. of Economic Behavior and Organization 65 (2008) 768.

[14] G. K. Rohde, J. M. Nichols, B. M. Dissinger, F. Bucholtz, Physica D 237 (2008) 619.

[15] N. Marwan, J. Kurths, Physica D 238 (2009) 1711.

[16] J. P. Zbilut, C. L. Webber, Phys. Lett. A 171 (1992) 199.

[17] C. Webber, J. P. Zbilut, J. of Applied Physiology 76 (1994) 965.

[18] M. Basseville, Signal Proc. 93 (2013) 621.

[19] J. V. Michalowicz, J. M. Nichols, F. Bucholtz, Phys. Lett. A 372 (2008) 7172.

[20] F. M. Birleanu, Ph.D. thesis, Université de Grenoble (2012).

[21] N. H. Packard, J. P. Crutchfield, J. D. Farmer, R. S. Shaw, Phys. Rev. Lett. 45 (1980) 712.

[22] F. Takens, Detecting strange attractors in turbulence., Vol. 898 of Dynamical systems and turbulence, Lecture Notes in Mathematics, Springer, Berlin, 1981.

[23] N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Physics Reports 438 (5-6) (2007) 237–329.

[24] R. A. Fisher, Biometrika 10 (1915) 507.

[25] H. E. Soper, A. W. Young, B. M. Cave, A. Lee, K. Pearson, Biometrika 11 (1917) 328.

[26] M. G. Kendall, A. Stuart, The advanced theory of statistics. Vols. I and II., Hafner, 1961.

[27] S. M. Ross, Introduction to probability and statistics for engineers and scientists, Academic Press, 2009.

[28] C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (1948) 379.

[29] S. Kullback, R. A. Leibler, The Annals of Mathematical Statistics 22 (1951) 79.

[30] M. M. Deza, E. Deza, Encyclopedia of distances, Springer, 2009.

[31] A. Cichocki, S. Amari, Entropy 12 (2010) 1532.

[32] D. Morales, L. Pardo, I. Vajda, J. of Stat. Planning and Inference 48 (1995) 347.

[33] M. Broniatowski, A. Keziou, J. of Multivariate Analysis 100 (2009) 16.

[34] H. Urkowitz, Proceedings of the IEEE 55 (1967) 523.

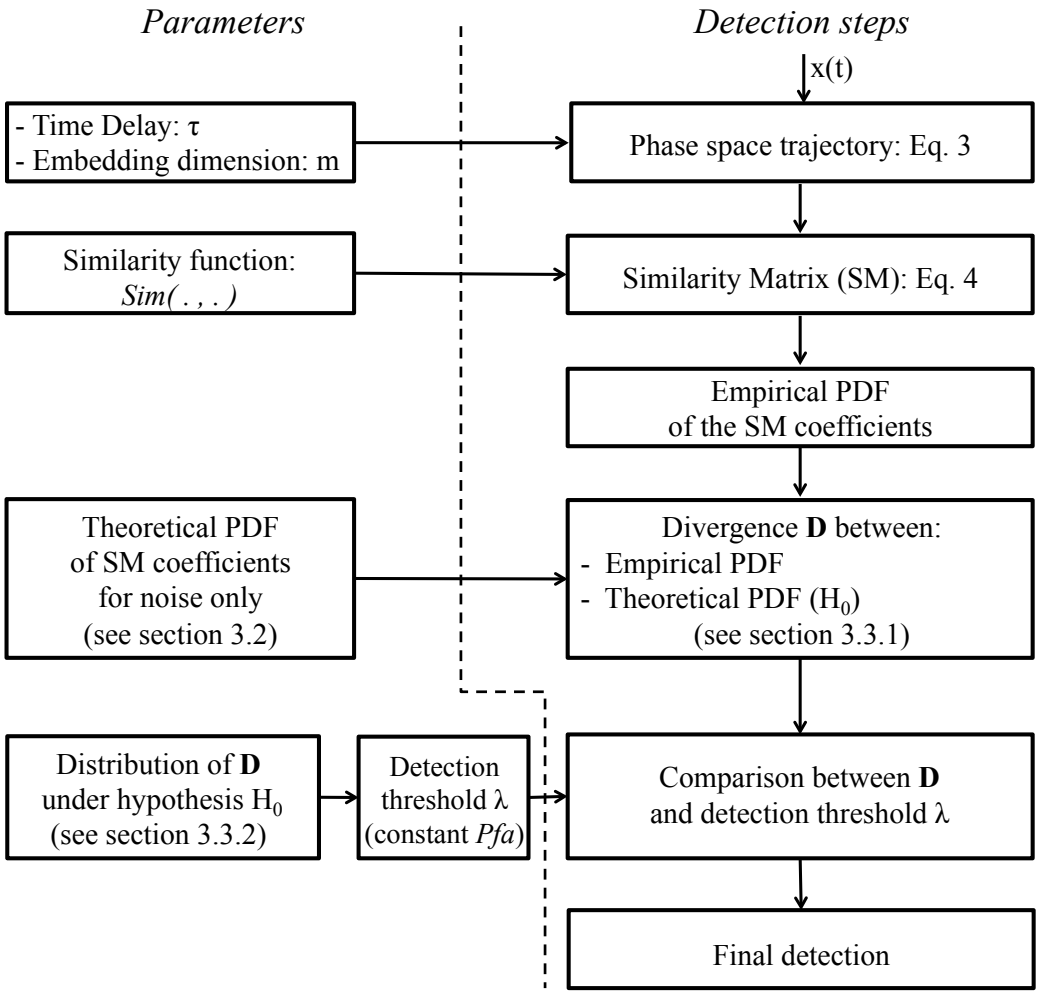

Figure 1: Flowchart of the detector system.

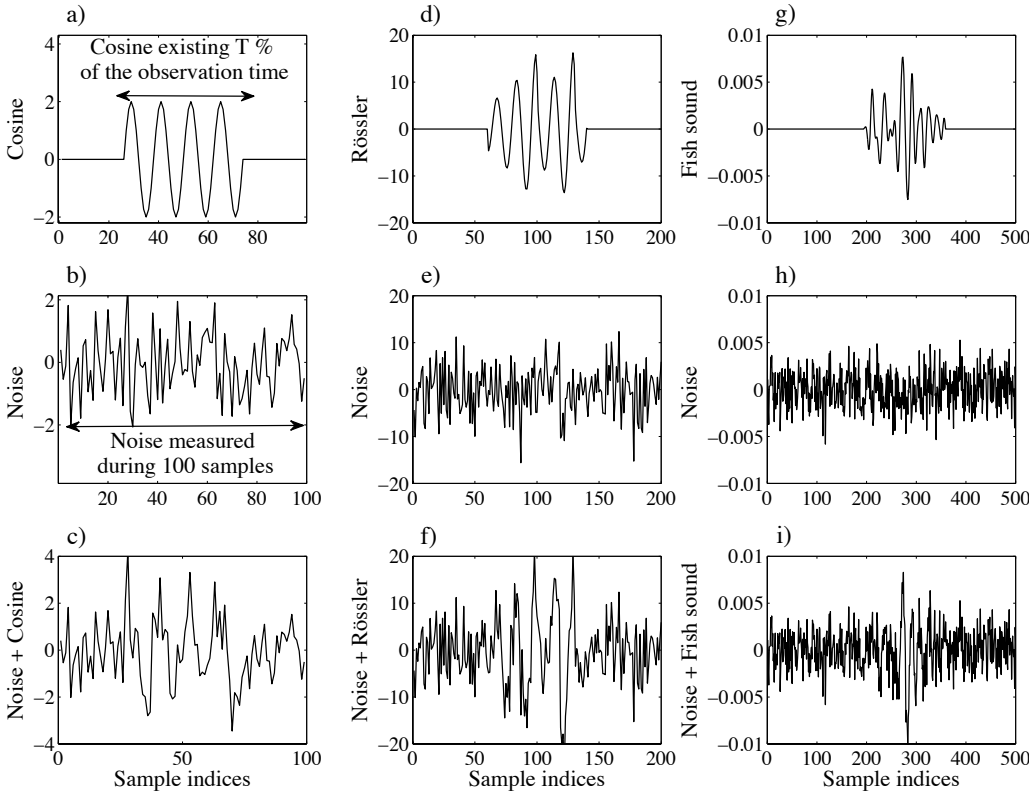

Figure 2: Under hypothesis H_1, the deterministic signal to detect that lasts $T\%$ of the observation time, $T \in \{10,\ 15,\ 20, ...,\ 80,\ 100\}$, is added to WGN. a-c) Example with the cosine function used in section 4.3; d-f) Example with a chaotic signal from a Rössler system used in section 4.4 ; g-i) Example with the waveform of the sound produced by a fish used in section 4.5. Figures c,f,i correspond to an SNR of 3 dB.

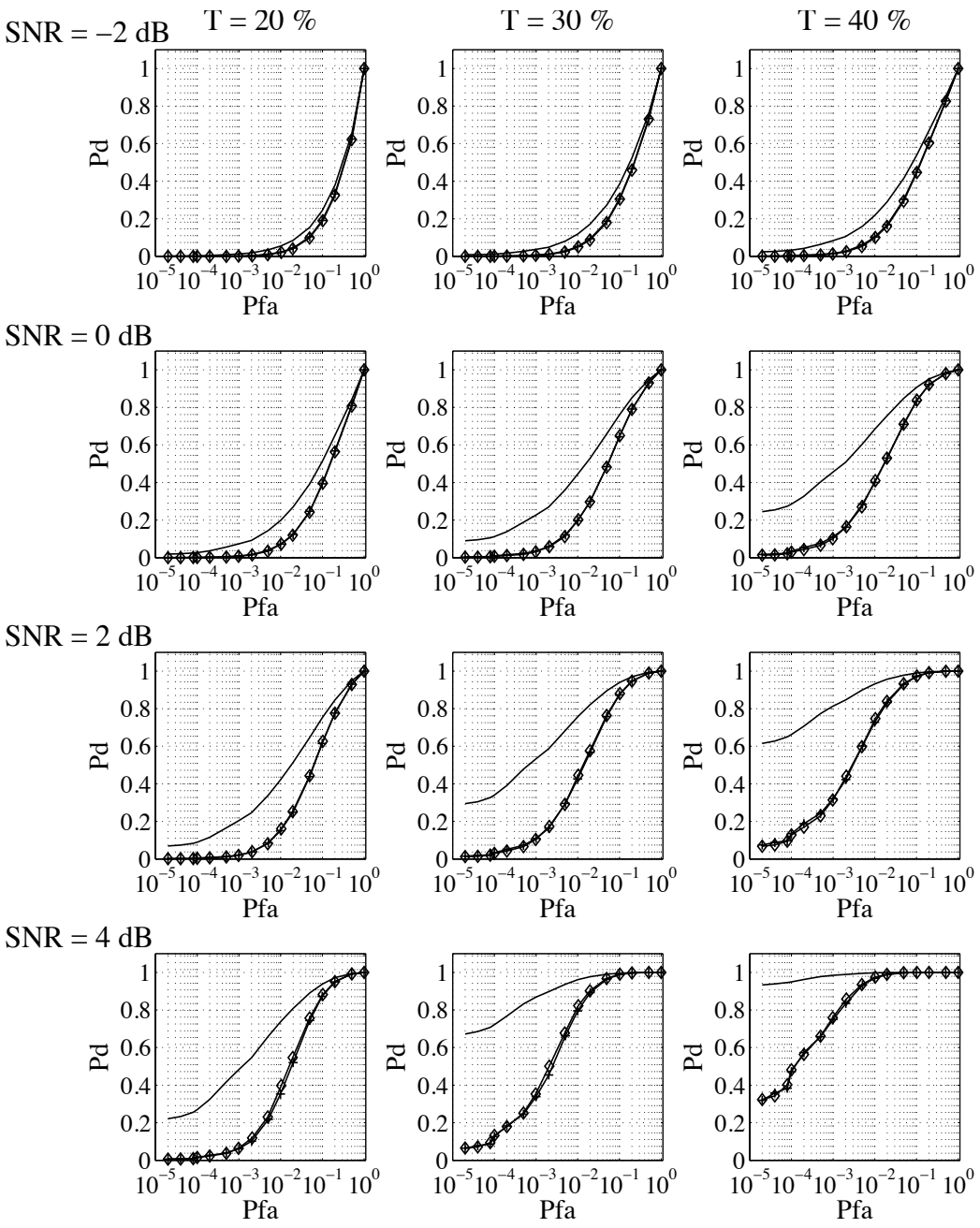

Figure 3: ROC curves based on the Euclidean norm as a function of the SNR (rows of panels), the length T of the cosine (columns of panels), and the Kullback-Leibler (plain line), Hellinger ($\diamond$), and Jensen-Shannon ($+$) divergence measures. The data for the Hellinger and Jensen-Shannon divergences are superimposed. The Kullback-Leibler divergence always outperforms the Hellinger and Jensen-Shannon divergences.

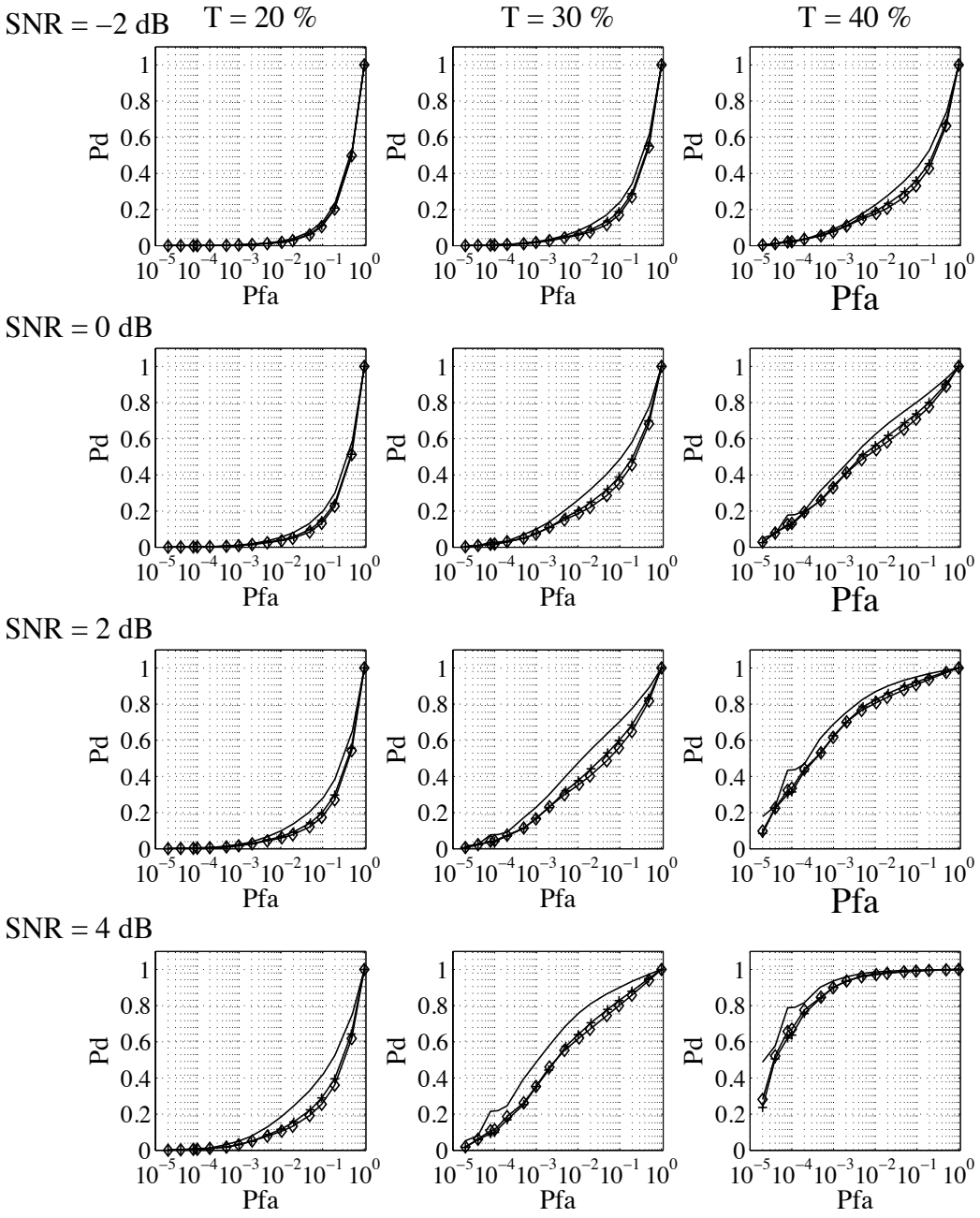

Figure 4: ROC curves based on Pearson's correlation coefficient as a function of the SNR (rows of panels), the length T of the cosine (columns of panels) and the Kullback-Leibler (plain line), Hellinger ($\diamond$), and Jensen-Shannon ($+$) divergence measures. The data for the Hellinger and Jensen-Shannon divergences are superimposed.

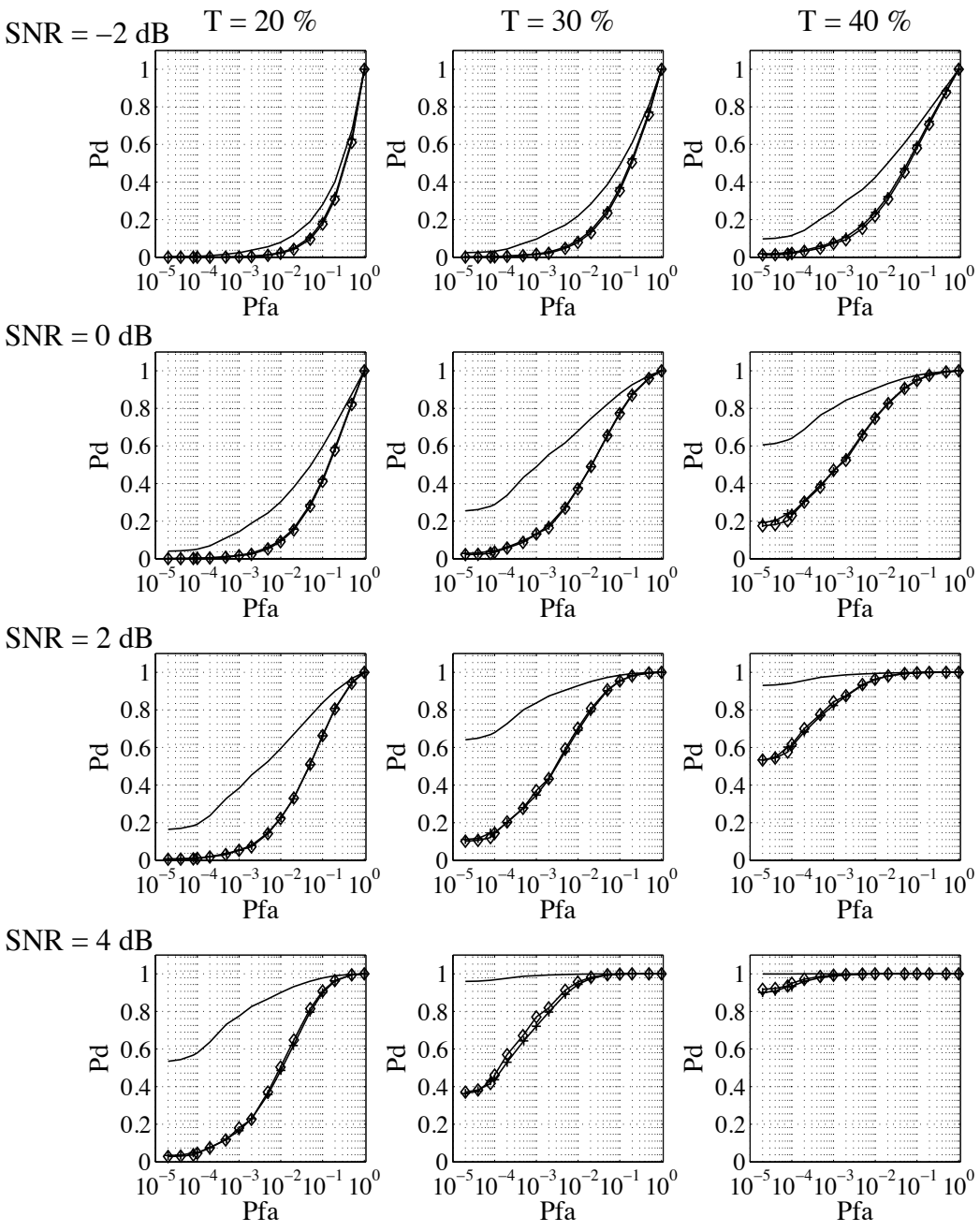

Figure 5: ROC curves based on the dot-product as a function of the SNR (rows of panels), the length T of the cosine (columns of panels) and the Kullback-Leibler (plain line), Hellinger ($\diamond$), and Jensen-Shannon ($+$) divergence measures. The data for the Hellinger and Jensen-Shannon divergences are superimposed. The Kullback-Leibler divergence always outperforms the Hellinger and Jensen-Shannon divergences.

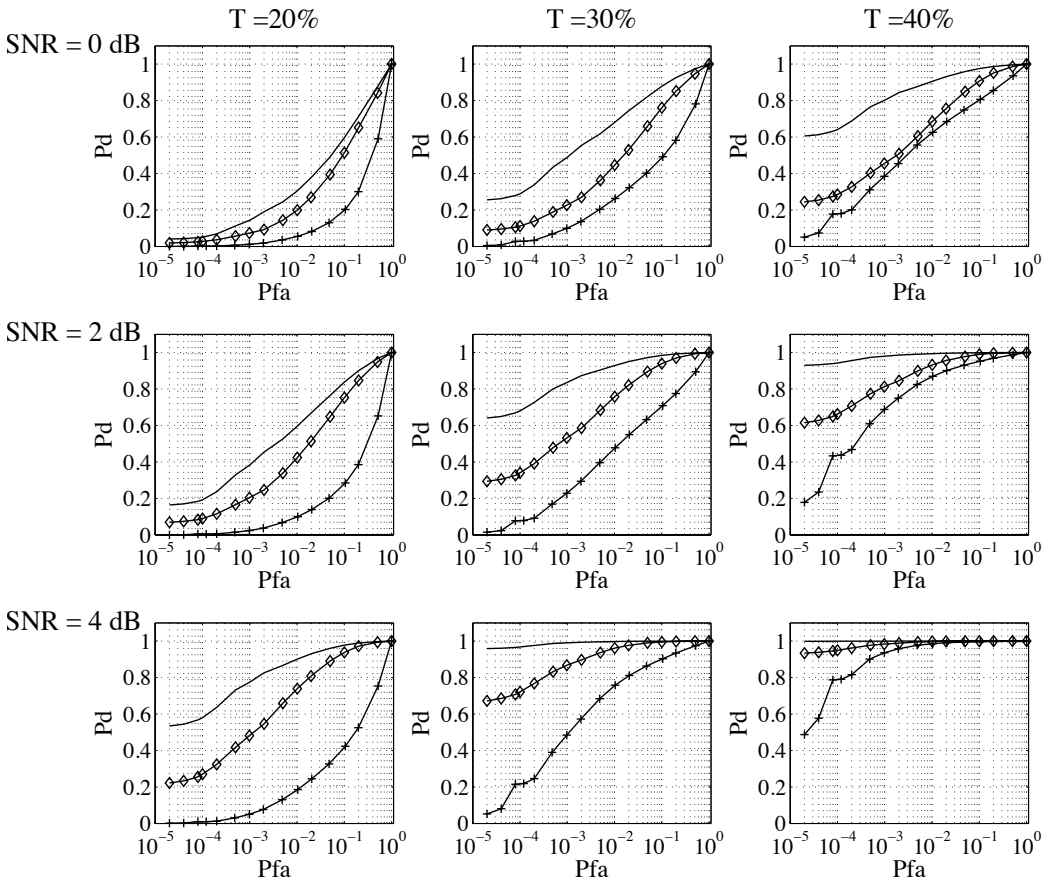

Figure 6: ROC curves of the detector as a function of the SNR (rows of panels) and the length T of the cosine (columns of panels), when the Kullback-Leibler divergence is associated with: the Euclidean norm ($\diamond$), Pearson's correlation coefficient ($+$), the dot-product (plain line). Combination of the dot-product with the Kullback-Leibler divergence always gives the best performance.

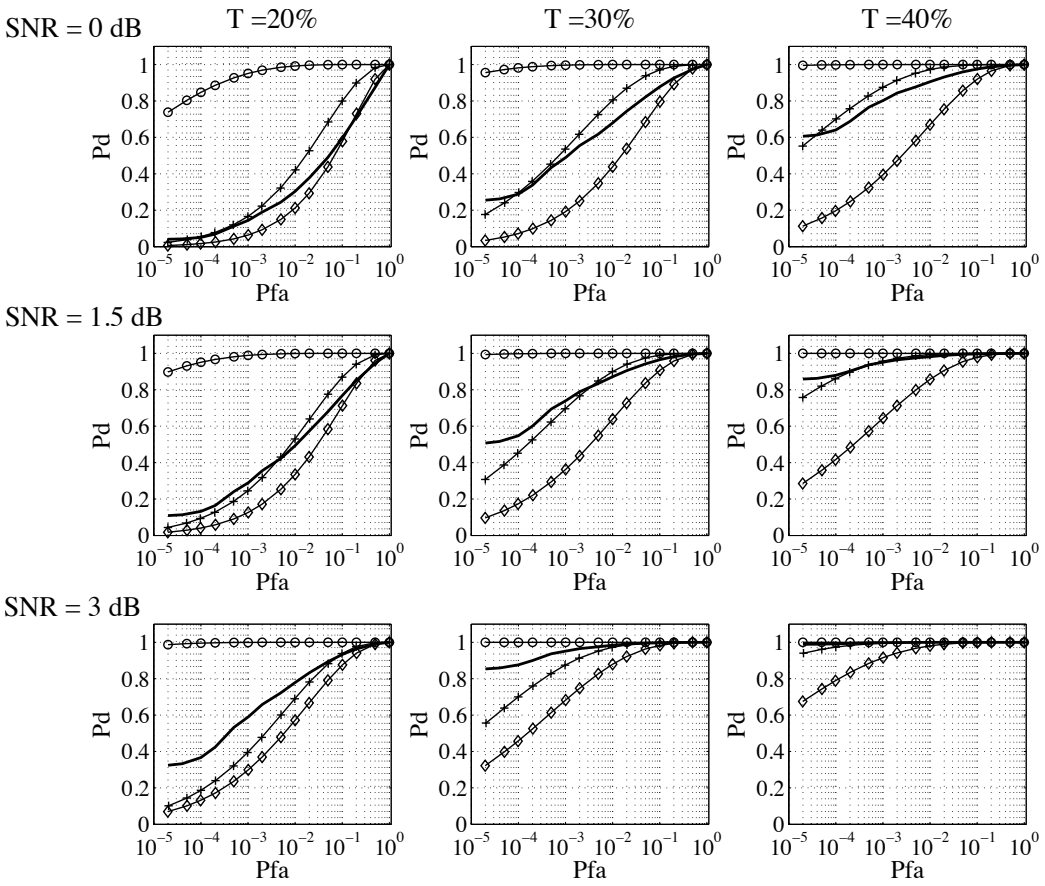

Figure 7: ROC curves of the proposed detector using {dot-product, Kullback-Leibler} (plain line) with the energy detector ($\diamond$), the sub-optimal filter ($+$), the optimal matched filter (o). The proposed detector always outperforms the energy detector, and globally it is as good as the sub-optimal filter detector.

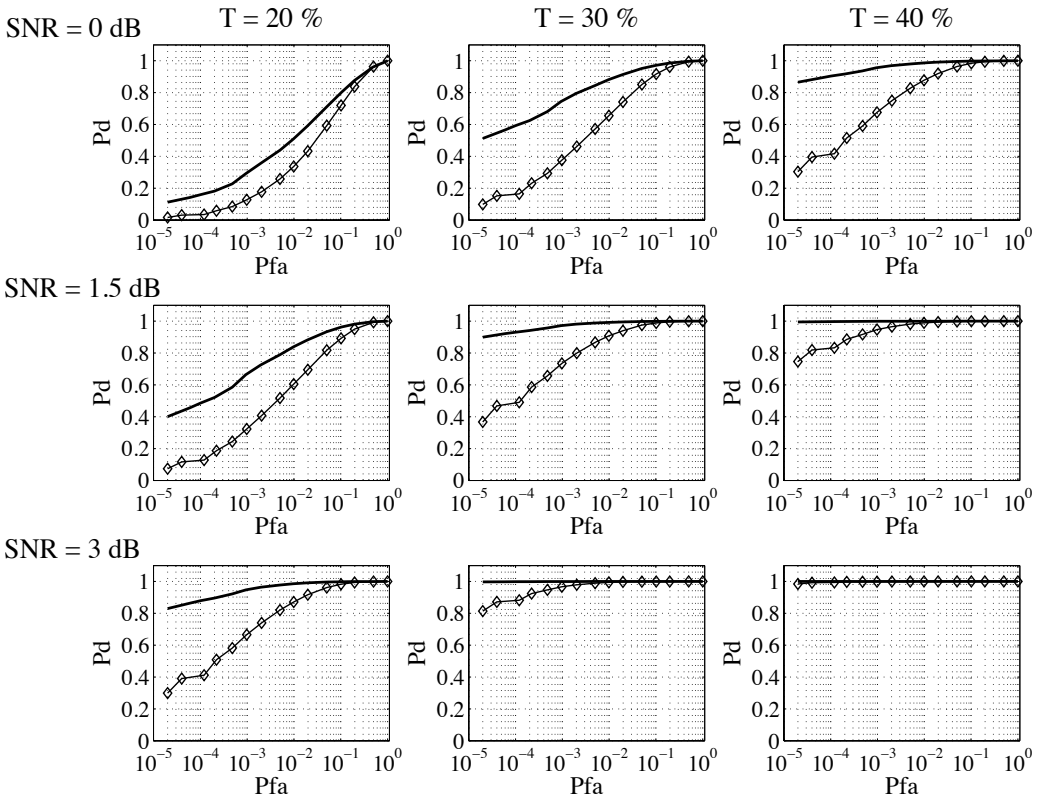

Figure 8: ROC curves of the proposed detector using $\{\tau = 1,\ m = 16$, dot-product, Kullback-Leibler$\}$ (plain line) with the energy detector ($\diamond$), when the deterministic signal to detect is a chaotic Rössler system. The proposed detector always outperforms the energy detector.

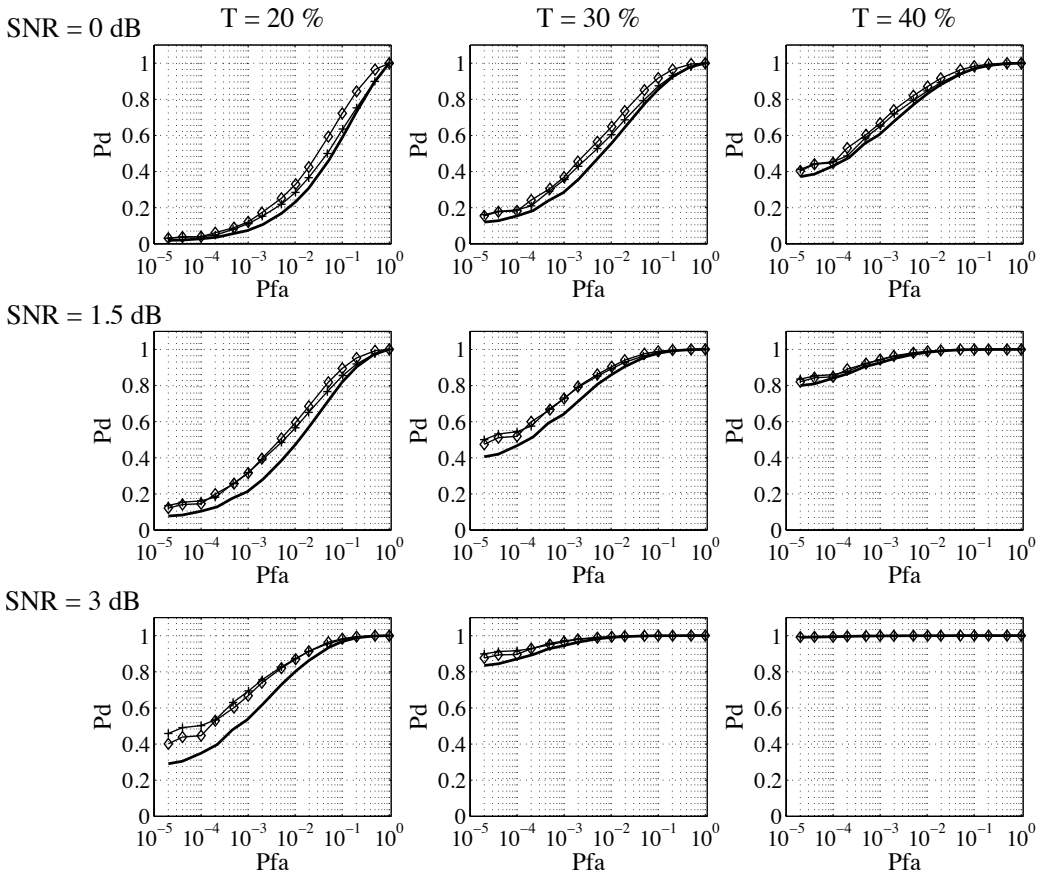

Figure 9: ROC curves of the proposed detector using $\{\tau = 3,\ m = 3,$ dot-product, Kullback-Leibler$\}$ (plain line), $\{\tau = 3,\ m = 3,$ Euclidean norm, Kullback-Leibler$\}$ (+) and the energy detector ($\diamond$), when the deterministic signal to detect is a chaotic Rössler system.

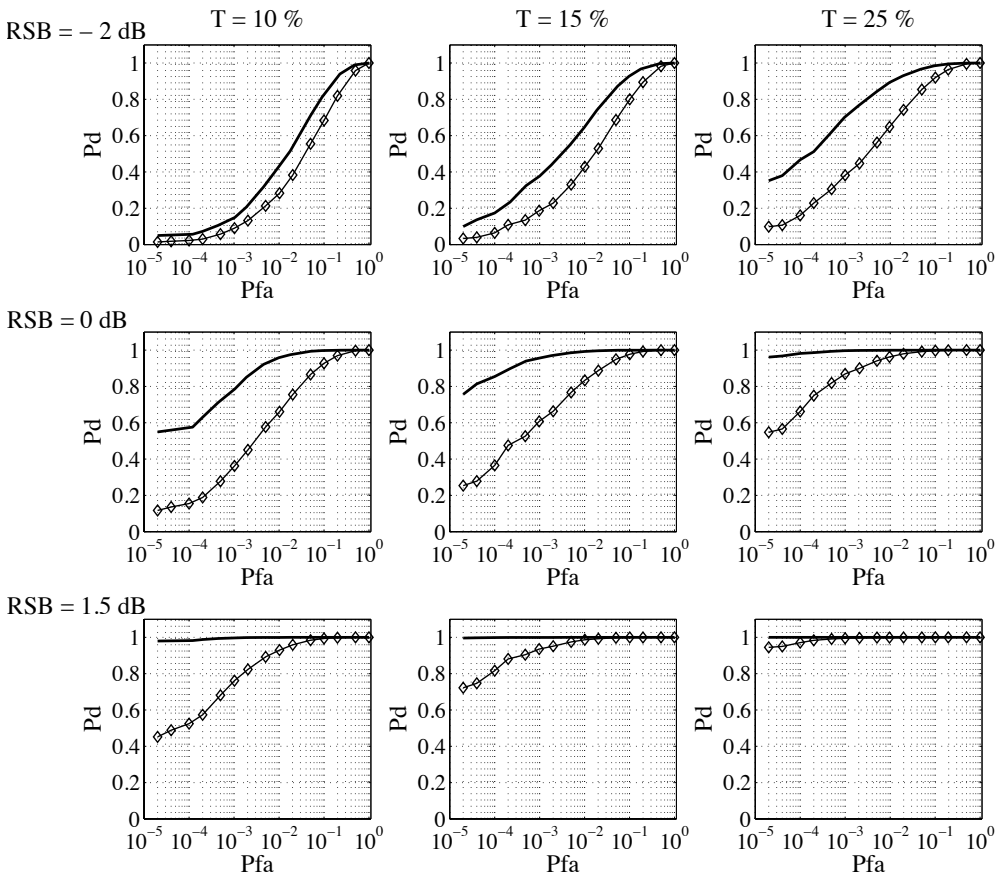

Figure 10: ROC curves of the proposed detector using $\{\tau = 1,\ m = 16$, dot-product, Kullback-Leibler$\}$ (plain line) with the energy detector ($\diamond$), when the deterministic signal to detect is a sound produced by a fish. The proposed detector always outperforms the energy detector.